

PROYECTO - ENTREGA 1

**DAVID MEJÍA CASTAÑO, 1007221842
ANDERSON VALENCIA BERMÚDEZ, 1000869230
DIEGO HERNANDO ARANGO RÍOS, 1042768156**

**INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL PARA LAS
CIENCIAS E
INGENIERÍAS / MODELOS Y SIMULACIÓN DE SISTEMAS I**

UNIVERSIDAD DE ANTIOQUIA

**DOCENTE
RAÚL RAMOS POLLÁN**

1. Problema predictivo a resolver

En el transcurso del programa académico, los estudiantes pasan por diferentes asignaturas que contienen diferentes niveles de dificultad. Existen algunas ocasiones en las que los estudiantes después de haber reprobado una asignatura un número determinado de veces, pierden la motivación de seguir adelante debido al incremento en la percepción de dificultad sobre la carrera y la disminución de su propia confianza.

La retención estudiantil es un desafío crítico para las instituciones de educación superior en todo el mundo. La capacidad de predecir qué estudiantes tienen un mayor riesgo de abandonar sus estudios antes de completarlos es esencial para que las instituciones puedan tomar medidas preventivas y ofrecer apoyo específico a los estudiantes en riesgo.

El objetivo principal de este problema es desarrollar un modelo de Machine Learning capaz de predecir si un estudiante dado continuará o abandonará sus estudios en la educación superior. Esto podría ayudar a las instituciones educativas a identificar a los estudiantes en riesgo y tomar medidas proactivas para retenerlos.

2. Dataset

El dataset que vamos a usar es el siguiente:

<https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention>. Este dataset cuenta con 4425 muestras y proporciona una serie de características que pueden utilizarse como predictores, como el rendimiento académico, la demografía de los estudiantes, la situación financiera, la participación en actividades extracurriculares, entre otros.

3. Métricas de desempeño requeridas (de machine learning y de negocio)

Para este caso, se pueden utilizar varias métricas de desempeño de machine learning y de negocio para evaluar el rendimiento del modelo. Algunas de estas pueden ser:

- Métrica de desempeño de machine learning:

Podemos utilizar algunas de las siguientes métricas:

Precisión(Accuracy): Esta métrica de precisión nos podría proporcionar información sobre la capacidad del modelo para clasificar correctamente a los estudiantes en función de la retención o el éxito universitario.

AUC-ROC (Área bajo la curva ROC): la medida de la capacidad del modelo para distinguir entre los estudiantes que abandonan o tienen éxito académico y los que no lo hacen.

- **Métrica de desempeño de negocio:**

Tasa de Retención Estudiantil: Esta es la métrica más fundamental y directa para evaluar el éxito de la retención de estudiantes. Se calcula como el porcentaje de estudiantes que continúan sus estudios en la institución durante un período específico, generalmente un año académico completo. El objetivo es aumentar esta tasa con el tiempo.

4. Primer criterio para el desempeño deseable en producción.

Se busca ofrecer un modelo predictivo que tenga la capacidad de identificar la retención estudiantil, que a través de las variables explicativas que resulten significativas, puedan tomarse medidas para dar respuesta de manera anticipada al problema de la permanencia estudiantil y esto conlleva que el desempeño deseable en producción nos arroje resultados porcentuales en cuáles estudios de nivel superior **tienen más de un 80%** de probabilidad culminar sus estudios sin novedades.

Este modelo debe analizar datos importante (demográficos y socioeconómicos) de los estudiantes y que se correlacionen con su rendimiento académico de esta manera se visualizan patrones y tendencias el cual logre predecir si continúa o no continua en el programa o área de interés que cursa actualmente.

La implementación de este modelo se realiza con el compromiso en que todo el plantel educativo como usuarios a la herramienta puedan introducir datos de manera sencilla y persuasiva; el cual ayude a optimizar tiempos y así hacerles un seguimiento detallado para todo aquel estudiante que esté en riesgo y de este modo se logre minimizar obstáculos que se les presenta al estudiante para continuar en la permanencia en la institución.

En resumen esta implementación del modelo debe ser sencilla, fácil de usar y escalable para garantizar que se pueda utilizar eficazmente en un entorno estudiantil que ayude a predecir qué estudiantes no tienen permanencia estudiantil **mayor al 20%**.